

AMENDMENTS TO THE CLAIMS

The following listing of claims replaces all prior versions, and listings, of claims in the application:

1. (Currently amended) A computerized method of generating a data mining model, the method comprising:

obtaining objectives for the data mining model;

automatically selecting a set of algorithms based on [[the]] objectives for the data mining model, the selecting including:

selecting a rule from a best practices database based on an objective for the data mining model; and

determining at least one of: an algorithm or a tuning parameter for an algorithm using the rule;

obtaining sample data;

creating a plurality of datasets from [[the]] sample data, each dataset including a unique subset of the sample data;

optimizing the set of algorithms using the plurality of datasets; and

generating the data mining model based on the optimized set of algorithms, wherein the data mining model mines data when executed.

2. (Currently amended) The method of claim 1, wherein the creating step includes:

shuffling the sample data, the shuffling including changing an order of entries in the sample data in a random fashion;

placing the shuffled sample data into a plurality of partitions, each partition including a unique subset of the shuffled sample data; and

including each partition in one of the plurality of datasets.

3. (Original) The method of claim 2, wherein the plurality of datasets includes a training dataset, a validation dataset, and a testing dataset.

4. (Currently amended) The method of claim 3, wherein the creating step further includes repeating the including step until to create multiple permutations of the plurality of datasets, wherein each partition is included in at least one the training dataset for at least one permutation.

5. (Canceled)

6. (Currently amended) The method of claim [[5]] 1, wherein the best practice is based on at least one of: research, data characteristics, and or user feedback.

7. (Currently amended) The method of claim 1, wherein the selecting step includes analyzing an attribute of the sample data, and wherein the set of algorithms is further selected based on the attribute.

8. (Currently amended) The method of claim 1, wherein the optimizing ~~step~~ includes:

applying the set of algorithms to the plurality of datasets; and

analyzing a set of results for the applying ~~step~~.

9. (Currently amended) The method of claim 8, wherein the optimizing ~~step~~ further includes:

adjusting at least one algorithm based on the set of results; and

applying the adjusted set of algorithms to the plurality of datasets.

10. (Currently amended) The method of claim 1, wherein the generating ~~step~~ includes translating the optimized set of algorithms into a set of standard query language (SQL) statements, and including the set of SQL statements in the data mining model.

11. (Original) The method of claim 1, further comprising storing the data mining model as a character large object (CLOB) in a database.

12. (Currently amended) A computerized method of generating a data mining model, the method comprising:

obtaining a set of algorithms;

automatically generating a plurality of datasets from sample data[[;]], wherein the automatically generating includes:

changing an order of entries in the sample data in a random fashion;

placing each of the entries into one of a plurality of partitions; and

creating the plurality of datasets based on the plurality of partitions, each dataset including at least one of the plurality of partitions;
optimizing the set of algorithms using the plurality of datasets, the optimizing including:

- applying the set of algorithms to the plurality of datasets;
- analyzing a set of results for the applying step;
- adjusting at least one algorithm based on the set of results; and
- applying the adjusted set of algorithms to the plurality of datasets; and

generating the data mining model based on the adjusted optimized set of algorithms,

wherein the data mining model includes a set of SQL statements.

13. (Canceled)

14. (Currently amended) The method of claim 12, wherein the obtaining step includes:

- obtaining objectives for the data mining model; and
- automatically selecting the set of algorithms based on the objectives.

15. (Currently amended) A system for generating a data mining model, the system comprising:

- a dataset system for automatically generating a plurality of datasets from sample data,
each dataset including a unique subset of the sample data;
- a rules system for obtaining automatically selecting a plurality of algorithms, wherein the
rules system selects a rule from a best practices database based on an objective for the data

mining model, and determines at least one of: an algorithm or a tuning parameter for an algorithm using the rule;

an optimization system for optimizing the set of algorithms using the plurality of datasets; and

a model system for generating the data mining model based on the optimized set of algorithms, wherein the data mining model includes a set of SQL statements.

16. (Original) The system of claim 15, further comprising a storage system for storing the data mining model in a database.

Claims 17-18 (Canceled)

19. (Currently amended) A program product stored on a recordable medium for generating a data mining model, which when executed comprises:

program code for automatically generating a plurality of datasets from sample data, each dataset including a unique subset of the sample data;

program code for automatically selecting a set of algorithms based on objectives for the data mining model, wherein the program code for automatically selecting selects a rule from a best practices database based on an objective for the data mining model, and determines at least one of: an algorithm or a tuning parameter for an algorithm using the rule;

program code for optimizing the set of algorithms using the plurality of datasets; and

program code for generating the data mining model based on the optimized set of algorithms, wherein the data mining model mines data when executed.

20. (Original) The program product of claim 19, further comprising program code for storing the data mining model as a character large object (CLOB) in a database.

21. (Previously presented) The program product of claim 19, wherein the program code for generating the data mining model includes program code for translating the optimized set of algorithms into a set of standard query language (SQL) statements, and including the set of SQL statements in the data mining model.

22. (Original) The program product of claim 19, wherein the program code for generating the plurality of datasets includes:

program code for shuffling the sample data;
program code for placing the shuffled sample data into a plurality of partitions; and
program code for including each partition in one of the plurality of datasets.

Claims 23-24 (Canceled)

25. (New) The method of claim 2, wherein the shuffling includes:

adding a fold data item to each entry;
generating a random number for the fold data item for each entry; and

resorting the entries based on the fold data items.

26. (New) The method of claim 12, wherein the plurality of datasets comprises a permutation of the partitions, the automatically generating further including repeating the creating to create a plurality of permutations of the partitions.

27. (New) The method of claim 26, wherein each permutation of the partitions includes a training dataset and each partition is included in the training dataset for at least one of the plurality of permutations, and wherein each applying includes training the set of algorithms using at least one training dataset.

28. (New) The method of claim 27, wherein each permutation of the partitions further includes a validation dataset and a testing dataset and each partition is included in the validation dataset for at least one of the plurality of permutations and the testing dataset for at least one of the plurality of permutations, and wherein the applying the adjusted set of algorithms further includes:
validating the adjusted set of algorithms using at least one validation dataset; and
testing the adjusted set of algorithms using at least one testing dataset.